

Time to say goodbye?

On the exclusion of solid compounds from the Swedish Academy Glossary (SAOL)

Sture Berg, Louise Holmer and Emma Sköldberg

Centre for Lexicography and Lexicology. Department of Swedish. University of Gothenburg

The Swedish Academy Glossary, SAOL (*short for Svenska Akademiens ordlista*) is a monolingual glossary, first published in 1874. The latest edition, SAOL13, was published in 2006 and the next edition, SAOL14, is planned for 2015.

This article concerns the revision of the lemma list in SAOL, with special focus on the exclusion of transparent solid compounds. There are about 88,000 solid compounds in the 13th edition of the Glossary, i.e. 70 % of the total number of lemmas (125,000). Since there are almost infinite possibilities of creating new words in Swedish, the printed Glossary obviously only includes a sample of the contemporary Swedish vocabulary.

*With improved lexicographic tools and an enlarged text corpus, the editors of SAOL14 have great possibilities of making more accurate decisions when including new solid compounds and excluding others from the lemma list. The discussion is above all based on the solid compounds including the noun *kalkyl* ('calculation', 'estimate', 'calculus').*

1. Introduction

The Swedish Academy Glossary, SAOL (short for Svenska Akademiens ordlista), has a long tradition in Sweden. It is a monolingual glossary, first published in 1874. Since then, a new edition has appeared about every tenth year. The latest, 13th edition, was published in 2006 and a CD-version of this edition, called SAOL Plus, was released in 2007. The next edition, SAOL14, is planned for 2015, with Professor Sven-Göran Malmgren as editor-in-chief. SAOL is financed by the Swedish Academy. Since 1984 and the preparation of the 11th edition, the Glossary has been compiled at the Department of Computational linguistics and later on at the Centre for Lexicography and Lexicology at the University of Gothenburg.

SAOL has a unique position in Sweden as it is widely used as a public reference work with regard to orthography and inflection. SAOL comprises about 125,000 lemmas, and although it is not a traditional dictionary, about one fifth of the lemmas are defined, commented on or syntactically exemplified. SAOL has a few equivalents in other countries, for example *Stafsetningarorðabókin* in Iceland, *Tanums store rettskrivningsordbok* in Norway, *Retskrivningsordboken* in Denmark and *Het Groene Boekje* in the Netherlands (see Vikør 2009 for a comparison of these works).

Since there are almost infinite possibilities of creating new words in Swedish, for example temporary solid compounds and derivatives, the printed Glossary obviously shows only a sample of the Swedish vocabulary. Thus, the editors have to make a selection of which lemmas to include and which to exclude.

Revision of the lemma list, such as including new words and excluding old ones, is of course a fundamental part of the editorial work before publishing any dictionary. Since SAOL has such a long tradition and is published so regularly, with the aim of representing contemporary Swedish, great effort has to be put into including new lemmas of current interest as well as excluding obsolete ones.

The aim of this article is to discuss lemma exclusion from the latest editions of the Glossary. The main focus is on the exclusion of solid compounds. Furthermore, we present and discuss

the lexicographic tools that we use nowadays for revision of the lemma list when preparing SAOL14.

2. Lemma list and lemma exclusion

In almost every theoretical work or handbook of lexicography, the size and the contents of the lemma list in existing or planned dictionaries are discussed (see e.g. Atkins & Rundell 2008:178–190, Svensén 2009:63–75). Such discussions have, of course, also preceded the appearance of different editions of SAOL. The number of lemmas has varied considerably during the 130 years of its existence, as shown in Table 1 (cf. Malmgren 2002).

Edition	Year	Editor-in-chief	Lemmas
1–5	1874–1883	F.A. Dahlgren	35,000
6	1889	F.A. Dahlgren	40,000
7	1900	Otto Hoppe	71,000
8	1923	Ebbe Tuneld	85,000
9	1950	Pelle Holm	155,000
10	1973	Gösta Mattsson	135,000
11	1986	Martin Gellerstam	115,000
12	1998	Martin Gellerstam	120,000
13	2006	Martin Gellerstam	125,000

Table 1. Editions of SAOL: year of publication, editors-in-chief and number of lemmas

As seen in Table 1, the first six editions have rather few lemmas, i.e. 35,000 – 40,000, compared to, for example, the 9th edition with no fewer than 155,000 lemmas. The large number of lemmas in this edition was justified by developments – and thus increased vocabulary – in, for example, welfare services, leisure time, technology and science. Also, more informal words were added (SAOL 9: Introduction). However, the new editor of SAOL 10 considered the lemma list too extensive, reducing it dramatically by 20 000 lemmas (see Mattsson 1974:79). Lemmas found to be obsolete or lemmas representing nonexistent institutions, etc., were excluded, but also many current words, especially transparent solid compounds (SAOL 10: Introduction). The next edition, SAOL 11, was also revised to a large extent and about 20 000 lemmas were excluded. This was due to the increased number of space-consuming definitions. In the 11th edition as well as in the 10th, it was primarily transparent solid compounds that were excluded (SAOL 11: Introduction). In the latest three editions, the number of lemmas has remained fairly stable at about 120,000 lemmas, and most likely the 14th edition will contain approximately the same number.

As already mentioned, SAOL has existed for 130 years, and if the lemma list in every edition is to reflect modern language use, it needs to be updated on a regular basis. Even between two editions of SAOL, the vocabulary of the Swedish language community will develop a great deal. Many new words appear, for example by borrowing from other languages, while others are used less. One reason for less frequent use is that the reality behind the word changes or disappears. For example, a *mjölkbod* ‘milk shop’ is no longer part of Swedish society so there is no need for this lemma in the Glossary. Further, some words develop negative connotations and are replaced by others (see Johannisson 1974:43–53; Mattsson 1974:81–88 for more examples).

Gellerstam (2009a:69) points to the problems of formulating a rule of thumb on how to exclude lemmas. Relevant factors are, according to Mattsson (1974:81), the meaning,

frequency and relative importance of the word, but he admits that elements of subjectivity are inevitable in the lemma exclusion. However, as Loman (1986) writes (our translation): “Most laymen probably do not think about how diffuse the border is between words accepted in SAOL and those not accepted, and about the agony the editor-in-chief has suffered when making selections for SAOL.”

Changes in the lemma list between two editions receive a great deal of attention in the press and from the public. The main reason for the interest in new lemmas may be the popular idea among some of its users that SAOL includes all “accepted” Swedish words and no other words. Or, to put it differently, if a word is not listed in the Glossary, it simply does not exist at all (Gellerstam 2009b:15). Some exclusions provoke negative reactions from users and are criticized in press reviews. For some reason, many people tend to respond negatively when a word is excluded, even if they have never used the word (Gellerstam 2009b:24–27). In such cases, the editors-in-chief of SAOL tend to point out that SAOL is a contemporary glossary, not a historical dictionary. Words excluded from the Glossary are not gone; many of them are accounted for in the comprehensive historical *Swedish Academy Dictionary* (SAOB). A complete list of the 5 500 lemmas excluded from the 12th edition is also available on the Internet as well as on the CD-version SAOL Plus.

3. Swedish solid compounds – in general and in SAOL

Many of the new words formed in Swedish develop from words and morphemes already established in the language, especially when it comes to new solid compounds (Svanlund 2009:15, 18). According to Malmgren (1994:32–43) and Teleman et al. (1999, 1:221), a solid compound is a word that can be split in at least two word-like parts, containing at least one core morpheme: *järn+väg* (‘railway’). The first part is called *modifier* and the second *head*. But many solid compounds are themselves part of other, more complex solid compounds: *järnväg+s+nät* (‘railway network’) and *järnväg+s+bro* (‘railway bridge’). Many solid compounds may also be combined with derivatives: *järnväg+s+förbindelse* (‘railway connection’) and *järnväg+s+korsning* (‘railway crossing’).

The elements making up a solid compound are above all nominal, adjectival or verbal, for example *barn+bok* (‘children’s book’), *skog+rik* (‘well-forested’), *hals+hugga* (‘behead, decapitate’), *privat+person* (‘private person’), *grön+gul* (‘greenish-yellow’), *stor+skratte* (‘roar with laughter’), *läs+rum* (‘reading-room’), *sitt+riktig* (‘...designed for sitting comfortably’) and *läs+träna* (‘train one’s reading’). The most frequent combination is noun + noun (Malmgren 1994:34). The inflection of the compound usually follows the last element: *bok* – *böcker* (‘book – books’); *barnbok* – *barnböcker* (‘children’s book – children’s books’) etc. (Teleman et al. 1999, 1:221).

Between the first and the last element in a solid compound there is sometimes a linking morpheme (Sw. *fogemorfem*). Especially when it comes to nominal first elements, the rules for linking morphemes are very complicated. For the extensive category of nouns ending in a consonant, it is hard to know whether such nouns require the linking morpheme *-s* or not when forming the first element in a solid compound. Nouns like *bil* (‘car’) and *djur* (‘animal’) are not followed by *-s*: *bil+bälte* (‘seat belt’) and *djur+park* (‘zoological park’), while *stad* (‘city’) and *år* (‘year’) are normally followed by *-s*: *stads+kärna* (‘town centre’) and *års+inkomst* (‘annual income’). Some nouns, like *skog* (‘forest’), may – depending on the second element – appear with or without *-s*: *skog+vaktare* (‘forest warden’) but

skog+s+brand ('forest fire'). There are also cases where the *s*-link is optional: *bord+(s)+duk* ('tablecloth').

Simple nouns ending in an unstressed vowel (often *-a* or *-e*) show other patterns (e.g. *lampa* 'lamp' and *pojke* 'boy'). When these words form the first element in a solid compound, they lose their final vowel (*lamp+skärm* 'lampshade' and *pojklag* 'boys' team'). Sometimes the linking morphemes *-o* or *-u* occur for historical reasons: *gata* 'street' but *gatu+korsning* 'crossing'; *kyrka* 'church' but *kyrko+gård* 'church yard' (Malmgren 1994:7, 36–39; cf. Teleman 1970). Learners of Swedish as a second language, but even many mother tongue speakers, may find it difficult to know the correct linking morpheme for each individual case (Malmgren 1994:7; 36).

Svanlund (2009:15–18, 20–22) points out that the meaning of solid compounds has often been an object of interest. According to a more mechanical compositional view, our mental lexicon does not normally store regularly formed solid compounds, since their meaning may be considered obvious from their components (e.g. *dator + bord = datorbord*, 'computer' + 'table' = 'computer table'). Today however, many linguists argue that this is not to be taken for granted; many solid compounds may be interpreted in a variety of ways. However, the probability of different interpretations varies a great deal, and language users are often capable of inferring the intended meaning, being guided in the right direction by the context.

In lexicography, a frequently recurring term concerning the semantic description of solid compounds is *transparency*. According to Svensén (2009:65), a solid compound is transparent to a native speaker if the meaning of the whole word is basically the sum of the meanings of its component parts. However, transparency is a gradual phenomenon – the meaning of a solid compound may be considered more or less obvious. The solid compounds not regarded as transparent are *lexicalized*. A case in point is *järnväg* ('railway'), mentioned earlier. The meaning of this compound is not possible to deduce from the elements *järn* ('iron') and *väg* ('way'). Another example is the noun *landgång* ('land' + 'path'), having two meanings, 'gangway' as well as 'long open sandwich'. None of the meanings is totally transparent. The latter, however, is without doubt lexicalized.

Judgements of transparency are regularly made by lexicographers. These judgements have consequences for the solid compounds which are included and defined in the dictionary, since the dictionary can not list all existing – or potential – solid compounds. Lexicalized solid compounds are generally listed, but the transparent ones are not. However, solid compounds may also be listed, even if they are transparent, due to their formal behaviour (cf. the above discussion on linking morphemes; Bergenholtz et al. 1997:258–259, Svensén 2009:64–65). It is, of course, a difficult task to determine the transparency of a solid compound when one already knows the meaning of the word. Lexicographers, therefore, will have to imagine how a more or less average language user reacts when first coming across the word (Svanlund 2009:17).

It is also of great importance for lexicographers to distinguish between solid compounds that are obviously casual word combinations from those starting to become established in the language. Svanlund (2009:191) points to the following factors that have to be considered when determining the degree of establishment: (i) the frequency of the word; (ii) if language users know the word and are familiar with its usage; (iii) the productivity of the word (i.e. if it is possible to form derivatives from it); (iv) 'age' of the word; (v) if the word is self-explanatory or not when used in a complete sentence; and, finally, (vi) writers' use of

metalinguistic signals like quotation marks when using the word. The first two factors are probably the ones that most clearly show to what degree the solid compound is established.

As already mentioned, SAOL comprises about 125,000 lemmas; about 88,000 of them are solid compounds. Incidentally, one of the longest words in today's SAOL is *realisationsvinstbeskattning* 'taxing of capital gain', a combination of three words. As said before, the Glossary only shows a sample of solid compounds, since there are almost infinite possibilities of creating them in Swedish. The solid compounds included in SAOL should, first of all, be well established in the ordinary language, hence – among other things – having high frequency in contemporary texts. The solid compounds included can be lexicalized or transparent, and it is the transparent ones that cause the most obvious selection problems (see below).

Previous editors-in-chief, like Tuneld, Mattsson and Gellerstam (preparing, respectively, SAOL8, SAOL10 and SAOL 11–13), have commented on the fact that they have excluded, in particular, solid compounds when reducing the lemma list (Allén 1986:253–259; Gellerstam 2009a:58, 69–70). Tuneld states that he has especially excluded "temporary" solid compounds. However, it must have been difficult to establish whether the solid compounds were indeed only temporary without recourse to electronic corpora, which today's lexicographers are heavily dependent upon.

4. Databases and lexicographic tools

As mentioned before, SAOL11, 12 and 13 have been edited in Gothenburg, where the 14th edition is also in preparation. Our knowledge about Swedish vocabulary in written text has increased over the years, much owing to computer-based corpora and advanced search functions. At the Centre, we have access to a corpus, mainly consisting of newspaper texts, which is an extension of the Swedish Language Bank (<<http://spraakbanken.gu.se/>>). In preparing SAOL11 and 12, the corpus was used only to a lesser extent, partly to obtain information about the use of certain inflectional forms, partly to decide which new words to include and which obsolete words to exclude. The working process at that time was very time-consuming. In preparing the publication of SAOL13, conditions were somewhat different, since the editorial staff had access not only to the corpus, but also to a specially designed morphological database called SMDB (Swedish morphological database). SMDB consisted of the 120,000 lemmas in SAOL12, with their full inflectional forms (for details, see Berg & Cederholm 2001, Berg 2009).

By running the content of SMDB against the corpus, it is possible to obtain the frequency of each lemma (all inflectional forms included). This provides information about how common the lemmas in SAOL are related to, in particular, newspaper texts. In this way, it is easy to discover what lemmas do not occur in the texts. In the earlier version of SMDB, the lemmas with no hits were detected automatically; the lemmas that occurred once or more had to be sought one by one. Finally, about 5,500 lemmas were excluded while producing the 13th edition, since they were not used in the corpus, which at that time (2005) contained about 100 million words. These words also had very few hits in other Swedish text databases and on the Internet. Today, the corpus contains about 250 million words. The question is: would the results from 2005 change by using a more extensive corpus? A random sample, 50 of the excluded words beginning in *ka-* and *sp-* (e.g. *kabbeleksblad* 'marsh marigold leaf' and *sparvbo* 'sparrow's nest'), shows that only two of these words occur – with only one hit each – in the enlarged corpus. The other 48 words are still not used. In view of this result, the

exclusions based on the limited corpus seem justified. However, it should be noticed that more than 10,000 other lemmas without hits in the corpus were *not* excluded due to their use in other Swedish text databases.

SMDB today consists of the 125,000 lemmas in SAOL13 with connection to the corpus. This databank forms the basis of SAOL14. At this point in time we also have the possibility of ranking the solid compounds, both those already included in the Glossary, and those which could be included. This ranking gives interesting insights and ideas when it comes to the next edition of the Glossary. It turns out that despite the exclusion of 5,500 words in 2005, about 10 % of the lemmas in SAOL13 are not used at all in the corpus. Around 5 % have one hit and 10 % have 2–4 hits. In other words, about 25 % of the lemmas in today's Glossary have fewer than five hits in the corpus.

In the following discussion, solid compounds containing *kalkyl* ('calculation', 'estimate', 'calculus') are exemplified. The noun *kalkyl* may function as first or last element in a compound. To begin with, *kalkyl-* forming the first element is discussed, but the main discussion concerns solid compounds with *-kalkyl* as last element.

In the Glossary, there are two lemmas starting with *kalkyl-*, namely *kalkylmässig* ('from the point of view of calculation') and *kalkylränta* ('interest calculated for costing purposes'). The adjective *kalkylmässig* has ten hits in our corpus and the noun *kalkylränta* six hits. In the corpus, 22 additional nouns are found, as shown in Table 2. The translations provided are not meant to be idiomatic but are intended to demonstrate the structure of the compounds.

Solid compounds	Frequency
<i>kalkylprogram</i> 'spreadsheet application'	19
<i>kalkylark</i> 'calculation sheet'	6
<i>kalkylblad</i> 'calculation sheet'	6
<i>kalkylunderlag</i> 'calculation data'	4
<i>kalkylarbete</i> 'calculation work'	3
<i>kalkylmetod</i> 'calculation method'	3
<i>kalkylförmåga</i> 'calculation ability'	2
<i>kalkylmiss</i> 'calculation miss'	2
<i>kalkylmodell</i> 'calculation model'	2
<i>kalkylchef</i> 'calculation head'	1
<i>kalkylexempel</i> 'calculation example'	1
<i>kalkylexercis</i> 'calculation drill'	1
<i>kalkylhastighet</i> 'calculation speed'	1
<i>kalkylkostnad</i> 'calculation costs'	1
<i>kalkylkunskap</i> 'calculation knowledge'	1
<i>kalkylmanipulation</i> 'calculation manipulation'	1
<i>kalkylmaskin</i> 'calculation machine'	1
<i>kalkylmetodik</i> 'calculation methodology'	1
<i>kalkylprincip</i> 'calculation principle'	1
<i>kalkylpris</i> 'calculation cost'	1
<i>kalkylsystem</i> 'calculation system'	1
<i>kalkyltid</i> 'calculation time'	1

Table 2. The solid compounds with *kalkyl-* in the corpus, not included in SAOL13

As shown by the examples in Table 2, the solid compounds with *kalkyl-* do not contain linking morphemes. Further, most of the solid compounds occur only once in the corpus. They may be regarded as temporary, and consequently of minor interest regarding lemma selection. The noun *kalkylprogram* ‘spreadsheet application’ is fairly frequent in the corpus; it has actually more hits than the two lemmas with *kalkyl-* in the Glossary. Hence, the word *kalkylprogram* could well be included in SAOL 14. In this case, as there are only two words with *kalkyl-* in the Glossary, both being relatively frequent, probably neither of these lemmas will be excluded. This would result in three words having *kalkyl-* as its first element in SAOL14.

Table 3 below lists the 13 solid compounds having *-kalkyl* as its last element in SAOL13. The words are sorted by frequency in the corpus.

Lemma	Frequency
<i>glädjekalkyl</i> ‘unrealistic cost estimate’	140
<i>kostnads kalkyl</i> ‘cost estimate’	91
<i>sannolikhetskalkyl</i> ‘probability calculation’	31
<i>lönsamhetskalkyl</i> ‘profitability calculation’	19
<i>felkalkyl</i> ‘miscalculation’	17
<i>snabbkalkyl</i> ‘quick calculation’	11
<i>integralkalkyl</i> ‘integral calculus’	6
<i>differentialkalkyl</i> ‘differential calculus’	4
<i>drift(s)kalkyl</i> ‘running/production calculation’	4
<i>avkastningskalkyl</i> ‘yield/profit calculation’	2
<i>efterkalkyl</i> ‘post calculation’	2
<i>överslagskalkyl</i> ‘rough calculation’	2
<i>förkalkyl</i> ‘preliminary calculation’	1

Table 3. The lemmas with *-kalkyl* (‘calculation’, ‘estimate’, ‘calculus’) in SAOL13

Obviously, many of the words ending in *-kalkyl* are frequently used in texts (e.g. *glädjekalkyl* ‘unrealistic cost estimate’, *kostnads kalkyl* ‘cost estimate’ and *sannolikhetskalkyl* ‘probability calculation’). However, at least four of the words in the list have very few hits (e.g. *avkastningskalkyl* ‘yield/profit calculation’, *efterkalkyl* ‘post calculation’, *överslagskalkyl* ‘rough calculation’ and *förkalkyl* ‘preliminary calculation’). Based on frequency, these words could be excluded from the Glossary. Two of the solid compounds listed differ from the others, namely the mathematical terms *integralkalkyl* ‘integral calculus’ and *differentialkalkyl* ‘differential calculus’. They have relatively low frequency in the corpus, due to the fact that they belong to more technical language. Based on frequency and the fact that they are technical terms, it would be possible to exclude them as well.

The solid compounds in Table 3 show different degrees of transparency. While *glädjekalkyl* (‘unrealistic cost estimate’) is lexicalized, *kostnads kalkyl* (‘cost estimate’) and *lönsamhetskalkyl* (‘profitability calculation’) are fairly transparent. The four words with the lowest frequencies already discussed, i.e. *avkastningskalkyl* (‘yield/profit calculation’) etc., could also be regarded as fairly transparent – and, in consequence, easy to understand; this further supports the idea of excluding them from the Glossary.

More than 50% of the lemmas do not have a linking morpheme. As to the rest, the most common linking morpheme is *-s*, as in *kostnads kalkyl*. In *drift(s)kalkyl* (‘running/production calculation’) the linking morpheme is optional. However, it is important to point out that the

linking morphemes are primarily illustrated in connection with the first element in the dictionary, i.e. *glädje-*, *kostnad-*, etc.

In the corpus, there are no fewer than 136 different compounds ending in *-kalkyl*, including the lemmas in the Glossary. Table 4 shows examples of solid compounds with *-kalkyl*, not included in the Glossary at present.

Solid compound	Frequency
<i>riskkalkyl</i> ‘risk calculation’	28
<i>boendekalkyl</i> ‘living estimate’	25
<i>bilkalkyl</i> ‘car estimate’	14
<i>investeringskalkyl</i> ‘investment estimate’	12
<i>boendekostnadskalkyl</i> ‘housing costs estimate’	11
<i>nyttokalkyl</i> ‘benefit/profit calculation’	10
<i>bolånekalkyl</i> ‘home loan estimate’	8
<i>totalkalkyl</i> ‘complete estimate’	8
<i>intäktskalkyl</i> ‘receipts estimate’	7
<i>förhandskalkyl</i> ‘beforehand estimate’	7
<i>brokalkyl</i> ‘bridge estimate’	6
<i>slutkalkyl</i> ‘final calculation’	6
<i>samhällskalkyl</i> ‘society calculation’	5
<i>lånkalkyl</i> ‘loan estimate’	4
<i>grovkalkyl</i> ‘rough calculation’	4
<i>energikalkyl</i> ‘energy calculation’	4
<i>byggkalkyl</i> ‘building/construction calculation’	3
<i>livskalkyl</i> ‘life calculation’	3
<i>långtidskalkyl</i> ‘long-term calculation’	3
<i>maktkalkyl</i> ‘power calculation’	3
<i>miljöskalkyl</i> ‘environment calculation’	3
<i>klimatekalkyl</i> ‘climate calculation’	3
<i>rimlighetskalkyl</i> ‘probability calculation’	3
<i>hushållskalkyl</i> ‘household estimate’	2
<i>pizzakalkyl</i> ‘pizza calculation’	2
<i>lyckokalkyl</i> ‘happiness calculation’	2
<i>fredskalkyl</i> ‘peace calculation’	2
<i>sparkalkyl</i> ‘saving calculation’	2
<i>baklängeskalkyl</i> ‘backwards calculation’	1
<i>besökskalkyl</i> ‘visit calculation’	1
<i>detaljalkyl</i> ‘detail calculation’	1
<i>gladkalkyl</i> ‘happy calculation’	1
<i>luftkalkyl</i> ‘air calculation’	1
<i>moralkalkyl</i> ‘moral calculation’	1
<i>skadeståndskalkyl</i> ‘damage calculation’	1

Table 4. Examples of solid compounds with *-kalkyl* in the corpus, not included in SAOL13

As shown in Table 4, a majority of the words are composed of two nouns (e.g. *riskkalkyl* ‘risk calculation’ and *investeringskalkyl* ‘investment estimate’) (cf. Malmgren and Teleman et al. above). Other combinations are also represented, e.g. *grovkalkyl* (‘rough calculation’) and *totalkalkyl* (‘complete estimate’), based on an adjective and a noun. Regarding linking

morphemes, there are also examples involving *-e* and *-o*, as in *bolån+e+kalkyl* ('home loan estimate') and *nytt+o+kalkyl* ('benefit/profit calculation').

With reference to the frequencies of the compounds, at least the top six words in Table 4 are very strong candidates for SAOL14. These words are *riskkalkyl* ('risk calculation'), *boendekalkyl* ('living estimate'), *bilkalkyl* ('car estimate'), *investeringskalkyl* ('investment estimate'), *boendekostnadskalkyl* ('housing costs estimate') and *nyttokalkyl* ('benefit/profit calculation'). By contrast, there are evidently a number of compounds with only one hit each in the corpus. Among them are *lyckokalkyl* ('happiness calculation') and *gladkalkyl* ('happy calculation'), two interesting, but not so far established, alternatives to the lexicalized *glädjekalkyl* ('unrealistic cost estimate') (see Svanlund 2009:112–116). Without context, it is difficult, or even impossible, to infer the meanings of the solid compounds in every single case, especially when the last element *-kalkyl* has another meaning than 'economic calculation', like in *maktkalkyl* ('power calculation') and *moralkalkyl* ('moral calculation') (cf. Svanlund above).

To sum up, in the Glossary there are at least six words – of totally 13 – ending in *-kalkyl* with low or very low frequency in modern texts. On the other hand, there are at least six solid compounds in the corpus which, at least according to frequency, are much more important. What to do in such cases?

5. Closing words

With the intention of restricting the number of lemmas in SAOL to about 120,000, as well as providing a sample of typical contemporary Swedish in the printed version of the Glossary, it is of vital importance to revise the lemma list by replacing uncommon solid compounds with more frequently used ones. As shown in this article, there are several possibilities of changing the lemma list in the Glossary radically, based on the frequency data provided by the Swedish morphological database (SMDB) of today. As about 70 % of the lemma list consists of solid compounds, the revision and replacement of lemmas will be very time-consuming. Such a revision would also most likely attract negative attention and even give offence to the Swedish public.

SMDB is a powerful lexicographic tool. However, it only provides frequency data, taking no other factors into consideration. The frequency of a solid compound in modern texts is, as shown, the most important factor in the case of SAOL, since this Glossary – contrary to many other dictionaries – includes a large number of transparent solid compounds. In other words, due to the fact that the Glossary incorporates not only well-established lexicalized solid compounds, but also transparent ones, the number of 'lemma candidates' increases enormously. As argued here, however, the lexicographer also has to pay attention to the formal behaviour and the semantics of the compounds. Furthermore, as the corpus connected to SMDB mainly consists of newspaper texts, it is also necessary to consult other kinds of texts before saying a permanent goodbye to low-frequency lemmas.

References

- Allén, S. (1986). 'Inför tredje seklet'. In Allén, S., B. Loman & B. Sigurd (eds.). *Svenska Akademien och svenska språket. Tre studier*. Stockholm. 233–273.
- Atkins, B.T.S. & M. Rundell (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Berg, S. (2009). 'Om ordböjning och SAOL Plus'. In Gellerstam, M. (ed.). *SAOL och tidens flykt. Några nedslag i ordlistans historia*. Stockholm. 139–165.
- Berg, S. & Y. Cederholm (2001). 'Att hålla på formerna. Om framväxten av Svensk morfologisk databas'. In *Gäller stam, suffix och ord. Festskrift till Martin Gellerstam den 15 oktober 2001*. Meijerbergs arkiv för svensk ordforskning 29. Göteborg. 58–69.
- Bergenholtz, H., I. Cantell, R. Vatvedt Fjeld, D. Gundersen, J. H. Jónsson & B. Svensén (1997). *Nordisk leksikografisk ordbok*. (Skrifter utgitt av Nordisk forening for leksikografi 4.) Oslo: Universitetsforlaget.
- Gellerstam, M. (2009a). 'SAOL i många upplagor'. In Gellerstam, M. (ed.). *SAOL och tidens flykt. Några nedslag i ordlistans historia*. Stockholm. 53–83.
- Gellerstam, M. (2009b). 'Vad är Svenska Akademiens ordlista?' In Gellerstam, M. (ed.). *SAOL och tidens flykt. Några nedslag i ordlistans historia*. Stockholm. 11–30.
- Johannisson, T. (1974). 'Ordlistans första nio upplagor – en historik'. In Johannisson, T. & G. Mattsson (eds.). *Svenska Akademiens ordlista under 100 år*. (Skrifter utgivna av Svenska språknämnden 55.) Stockholm. 1–53.
- Loman, B. 1986. 'Ordboken, ordlistan och några andra av Svenska Akademiens språkliga insatser under 1900-talet'. In Allén, S., B. Loman & B. Sigurd (eds.). *Svenska Akademien och svenska språket. Tre studier*. Stockholm. 143–231.
- Malmgren, S-G. (1994). *Svensk lexikologi. Ord, ordbildning, ordböcker och orddatabaser*. Lund: Sudentlitteratur.
- Malmgren, S-G. 2002. 'Normering i Svenska Akademiens ordlista 1874–1950: principer och resultat'. In *LexicoNordica* 9. 5–20.
- Mattsson, G. 1974. 'Ordlistans tionde upplaga – riktlinjer och kommentarer'. In Johannisson, T. & G. Mattsson (eds.). *Svenska Akademiens ordlista under 100 år*. (Skrifter utgivna av Svenska språknämnden 55.) Stockholm. 55–116.
- SAOB = *Svenska Akademiens ordbok*, <http://www.saob.se/> [Access date 28 Febr. 2010].
- SAOL1–13 = *Svenska Akademiens ordlista*. Edition 1–13. Stockholm: Norstedts.
- Språkbanken, <http://spraakbanken.gu.se/> [Access date: 28 Febr. 2010].
- Svanlund, J. (2009). *Lexikal etablering. En korpusundersökning av hur nya sammansättningar konventionaliserats och får sin betydelse*. (Acta Universitatis Stockholmiensis. Stockholm Studies in Scandinavian Philology, New series 52.) Stockholm: Institutionen för nordiska språk, Stockholms universitet.
- Svensén, B. (2009). *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Teleman, U. (1970). *Om svenska ord*. Lund: Gleerups.
- Teleman, U., S. Hellberg & E. Andersson (1999). *Svenska Akademiens grammatik*. Part 1. Stockholm: Norstedts Ordbok.
- Vikør, L.S. (2009). 'Lexicography and language planning in Scandinavia and the Netherlands'. In: Nielsen, S. & S. Tarp (eds.). *Lexicography in the 21st Century. In honour of Henning Bergenholtz*. Amsterdam/Philadelphia. 123–143.